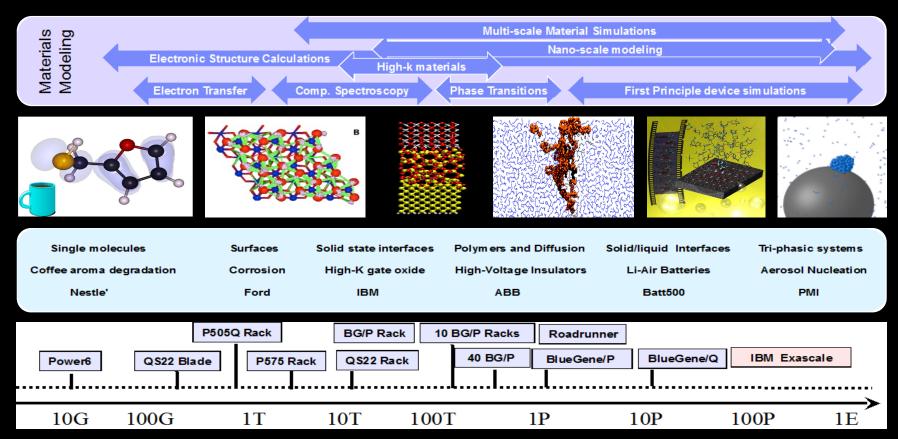
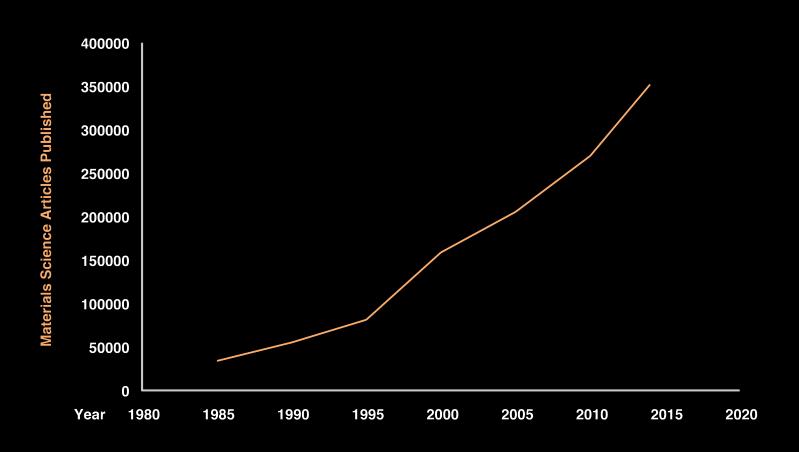
Machine Learning and Artificial Intelligence: applications to Materials Science for customized solutions

July 29, 2019 Materials Science in Space Workshop (ISSR&D)

HPC Enabled Simulations of Realistic Models



Unstructured Data Deluge in Peer Review Publications







C14ExOc2 256.283 Probably the commonset aldobioussuic said present as a structural unit in plant gums. Iso'l. from partial acid hydrolysates from the following plants: black worth cranophyllo, ugg plans (Fromus domestica) mar perrica). Anopeixus fertifelia (gum ghattiji. Eigilia srobnides, Afrangle jumine fate. Forms and Chevisio upp. Also incl. from hydrolysmus of maritime pine (Pour pinante) hemicalistics and wheat street Mp 118-119" (fiphose), $[x]_{0}$ +11.6 \rightarrow -8.6

Cr sol: $[x]_{\Omega} + 2 (H_2\Omega)$. Re sol: $[x]_{\Omega} - 3 (H_2\Omega)$.

Meaner CoHoOo 77031 $Mp.119^{\circ}, \|u\|_{D} = 0.(H_{2}\Omega).$ House de Me ester

C₂₅H₄₀O₂₄ 664.57 Mp 202-207: [u]₀ -17.5 (CHCl₄). f-Mr. 6-O-(#-O-Mala) f-o-planess 11006-61-2

Colikotto 77031 Structural unit in plant game, e.g. Preamaic infollow. Commissione provide [3]to -1 (H₂O).

n-Pyranou-form [52554-59-7] $[x]_{\Omega} + 2\{H_{2}\Omega\}.$ 1.2.3.4-Di-O-isomorphisms, 2.3.4-mi-to Me sener [15986-41-3] C₂₀H₂₀O₃₇ 576:55 Mp 114-115° [a]₀ =65 (c, 4.5 in CHCl₂) Me glysside, here-Me, Me etter C₂₆H₂₆O₁₂ 468.497 [10]0 +42 (CHICL).

E-Prospose-from (52554-60-6) Me glyonide, henryde, Me ono CaHaOa 616.56

Mp 140°. [a]p, -54 (CHCL): Me glassife bere-Me. Me ester [22854 Mp 86-99°. [a]o -25 (CHICL).

[1911-00-T. 5106-09-0] Hetaltina R.D. et al. J Red Clean. 1996.

Challed, W.F. et al., J. Red. Chees., 1998, 224, 207 (deal) Jackson, J. et al., J.C.N., 1940, 74 (a-Mr-per desardle Me ester (tot) August, G.O. et al., F.C.X., 1993, 1168, 1861,

3001 (nod) Muldanger, S. et al., J.4 C.S., 1998, 80, 2516

Sens, JK N et al., Gas J. Chem., 1961, 98, 162 Bales, R.W. et al., Olivensysberides, Personnes

Prent Lendon, 1903, 4, 13d (over); De K.K. et al., Carbolysk, Res., 1997, 4, 177 (oyali, disaproprilden brids Mecalis); Peant, C. et al., Clien, Zeyrit, 1970, 22, 17 Di Folio, Cl., et al., Carbolado Ros., 1987, 99.

cadesine, BAN, INN, USAN carbenowide SCLARCA-Riberide Al-



Accumulates in the culture medium of E. coli under sulforumide statis. Marsel by Ravillar passifier and Ravillas subrilis. nourtes (E10H). Sel. H₂O. Mp 215-216° disc $\{u_{11}^{(i)} -62.4 \text{ (s. } 0.5 \text{ in } H_2O_3 \text{ Log } P -2.97 \text{ (saks).}$

Piones [26418-25-5] Mp 146-147". 5'-Phophete: [1811-94-5] CaH₂₁N₂O₂P 338.213

Produced by the action of brown's years on the silvenile. The 5'-phosphate and its Nucleotide ignorhood in the demons himsynth. of other purine nucleotides. Converted by Bratton-Marskell reagent to a purple dye with λ_{max} 540 ran, λ_{max} 268 (c. 12880) (EL/O).

2, F, 5'-10'-4c' [33254-21-7] C₁₀R₁₀N₂O₂ 384.345 Mp of (88-80', 130-131'), [ag] -38.5

6c, I.0 in Br06h. N.Z.J. S.-Tone-Ac (17885-74-8) Cost. (E0040: Mg 208-209", D07" +2.9

6c. 0.51 in CBCl₄). 2,1'-O-laprophism, 5'-Ac Cullin Nath 140.135 Ma 169-169

2-Decay: Cayst. (MuOH). Mp 177-1781

Cryst. (MsO6i aq.). Mp 214-216'.

Caresdony, C.R. et al., J. Et al. Cham., 1956, 238, 411 (623 (see)) Bancore, LH, et al., J.C.S.(C)., 2007, 1888 (synth, plosphate)

DX Pol., 1967, 5 555 591; C4, 68, 28656q (exc); DX, Pol., 1968, 1 156 991; C4, 68, 28656q (exc); (copropriator Ac)
Pennin, R.F. et al., J.O.C., 1971, Mr. 1964

Pennics, R.P. et al., J. Hot. Chem., 2012, 9, 623 Jonesia, GA et al., JGC ., 1994, 29, 3651

Advanta D.A. et al., detections R. 1979, 98 724 (cryst struct)
Salvina, R.L. et al., J. Biol. Chem., 2162, 287, 19176 (motal) From FR et al., 2010, 1985, 98, 207 (seeds on

Midden, K. et al. Grahesay, Roy. 1919, 25 47 (ree) Incimale The Exten Pharmacopools, 19th ole, Planeacutical Press, 1965, 1926 Minakera, N. et al., Glore, Planes Bull., 1966, 44, 288 (7 sloves: Fulcocy) Estepana, N. et al. Speklent, 2001, 2691-2602 (profit, port cur)

4.6-Dideoxy-4-[4.5.6-trilipdeoxy-3-(kydeoxymethyl)-2-cyclobroom-1-yl jami no)-w-gheengemanyb-(1-4)-w-gheen gemanyb-(1-4)-w-gheens NCL Anyhonatio J. Gheeshay. Bay g 5421.

CallaNO₁₁ 645.611 A component of the Applicatetic complex See also Application, A-475. Incl. from Acting loss up. A potent inhibitor of u-glacesidases and excharases. Used for treatment of diabetes, leverlyidaemic and obsity Lounched in 1990. Amorph. Sol. H₂O; fairly sol. MuOSE: poorly tol. B2OH, houses: $|a|_D + 147.2$ (H_2O_3 , $|a|_D^{10} + 171.3$. ► LZ7153000

7-O-Phonbon C21H41NO21P 724.583

the Ast. 1975, 2 half 782; Cal. \$8, 587214; Salessall, DDD of al., Natural consultation.

Acaciabiuronic acid -

[7264-19-9]

molecular formula



Entry name

Acaciabiuronic acid

6-O-B-D-Glucopyranuronosyl-D-galactose. 9CI. 8CI

COOH

C12H20O12 356.283

Probably the commonest aldobiou acid present as a structural unit in plant gums. Isol. from partial acid hydrolysates from the following plants; black wattle (Acacia mollissima), Acacia senegal, Acacia pycnantha, Acacia karroo, Acacia

cyanophylla, egg plum (Prunus domestica).

Bailey, R.W. e. Press, Lond De, K.K. et al

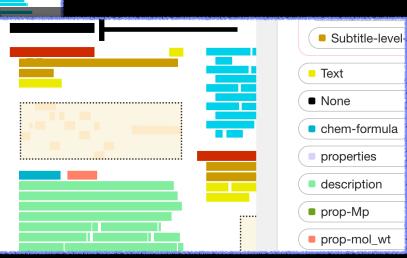
(isol)

(synth, diiso Peciar, C. et al (config. pmi

molecular weight

PDF Parser

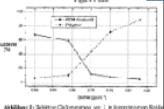
Handbook of Carbohydrates



Document Understanding Workflow

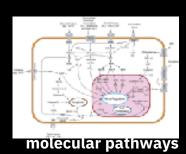


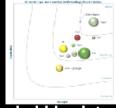
Image classification



flow charts

line and scatter plots





bubble plots

Major Ne	w York Bridges	
Bridge	Designer	Lengtl
Brooklyn	J. A. Roebling	1595
Manhattan	G. Lindenthal	1470
Williamsburg	L. L. Buck	1600
Queensborough	Palmer & Hambostel	1182
Triboragh	O. H. Ameraca	1380 383
Brons Whitestone	O. H. American	2300
Throgs Neek	O. H. Amexenn	1800
George Washington	O. H. Amexann	3500

scanned tables

Tables understanding

Table 4. Selected Interest Rates								
	2002				Week Ending			
Instruments	Apr 15		Apr 17			Apr 19	Apr 12	Mar 2002
Federal funds (effective)	1.84	1.68	1.73	1.71	1.67	1.78	1.71	1.73
Nonfinancial commercial paper								
1-month	1.76	1.74	1.74	1.74	1.72	1.74	1.77	1.78
2-month	1.78	1.75	1.80	1.75	1.75	1.77	1.78	1.82
3-month	1.82	1.82	1.83	1.78	1.81	1.81	1.82	1.86
Financial commercial paper								
1-month	1.76	1.77	1.76	1.75	1.74	1.76	1.77	1.80
2-month	1.80	1.80	1.79	1.78	1.77	1.79	1.79	1.82
3-month	1.83	1.84	1.83	1.83	1.81	1.83	1.84	1.87

Text analytics NLP

he composition and crystalline structure. When the cation M has a le a higher screening constant for an unoccupied orbital), the Li electrochemical stability against metallic Li. The calculation result studies in terms of the inertness of LizLa 272012 and L it cells the linkage of MO6 octahedrons also strongly influences instance, in spite of stability of the garnet-type LisLasTasO15 pe La_{1/3}TaO₃ with the same constituent elements is reactive in rnersharing MO₆ octahedral network. The Li + conductivity can als

Semantic analysis: unsupervised generic extraction

text ** "At ambient temperatures, tungsten is not attacked by hydrochloric, sulfuric nitric, and hydrofluoric acids at any concentration, or by aqua regia "



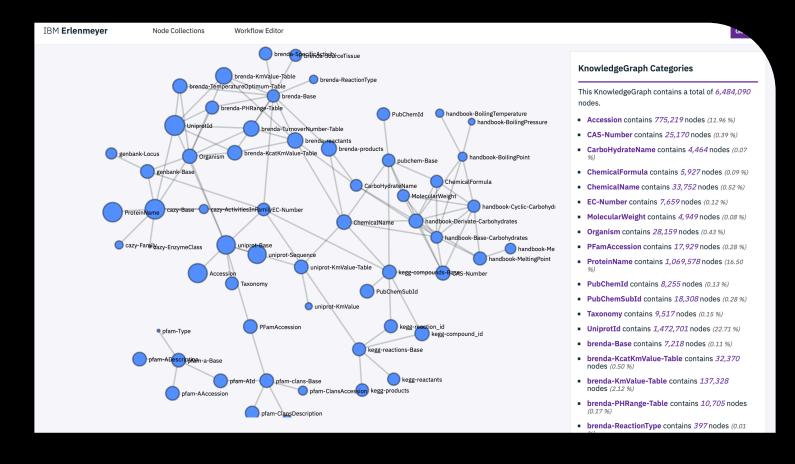
REL GENERIC TRIPLETS

```
tungsten // to_be__not__attack ambient__temperature // hydrochloric acid
tungsten // to be not attack ambient temperature // sulfuric acid
tungsten // to be not attack ambient temperature // nitric acid
tungsten // to_be__not__attack ambient__temperature // hydrofluoric__acid
tungsten // to be not attack ambient temperature // aqua regia
```

ADDITIONAL TRIPLETS

hydrochloric // HasType // chemical substance, chemical compound nitric // HasType // chemical substance, chemical compound hydrofluoric // HasType // chemical substance, chemical compound

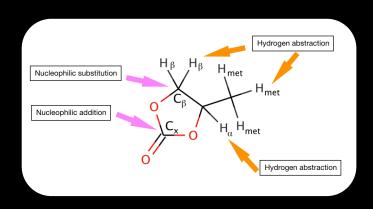
Domain Knowledge Graph



A real case: base-line

The Propylene Carbonate (PC) may degrade in different possible ways.

In 2015, we addressed this problem by doing extensive and CPU-intensive DFT calculations.



Battery 500

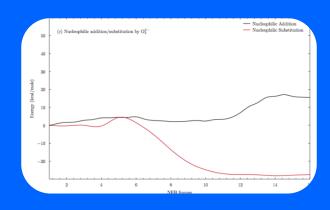
Organic liquid electrolyte chemical stability issue



Approximately 2 months to complete the analysis of the possible degradation pathways.

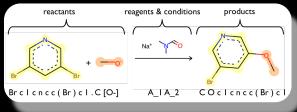
~ 5 Mio CPU-hours (BGQ) to generate reactive profiles (NEB) + 100 man-hours for analysis and simulation assistance

An example of an NEB outcome From reactants to products (PC degraded by peroxide)



A real case: Data and AI

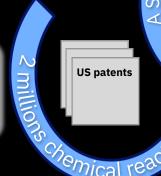
Forward Chemical reaction prediction



SMILES

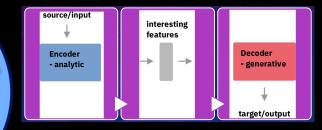
to

SMILES

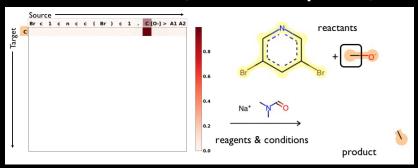


SMILES to SMILES prediction using sequence-2-

sequence models



Predictions in short (how a Se2Seq works)



Schwaller et. al. Chem. Sci., 2018, 9, 6091-6098

Performance

USPTO*		S2S	WLDN	ELECTRO	GTPN	WLDN5	our work
		[13]	[28]	[30]	[31]	[29]	
_MIT	separated	80.3	79.6		82.4	85.6	90.4
_MIT	mixed		74				88.6

Currently, best method in FORWARD chemical reaction prediction

Freely available for everyone:

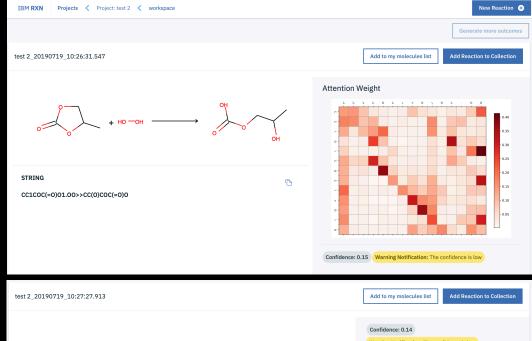
http://www.research.ibm.com/ai4chemistry

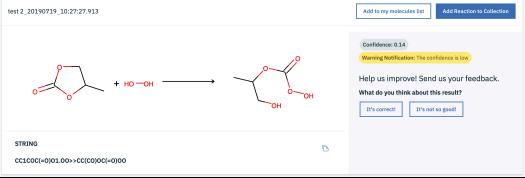


Released on August 19th at ACS Boston. In 10 months (Jun. 2019):

- 6300 users
- 48000 chemical reactions predictions

A real case: AI benefits





< 6 seconds prediction time for possible degradation products

Correct prediction order (Substitution more favorable than Addition)

Screening 30 different electrolytes using DFT took approximately 6 months (or **15 Mio seconds**). Using the AI model (RXN) for an initial screening this would have taken **300 seconds**. Accelerated by **50,000**.

Thank you.

Questions?

